Translation of a minigene in the 5′ leader sequence of the enterohaemorrhagic Escherichia coli LEE1 transcription unit affects expression of the neighbouring downstream gene

Islam, Md; Shaw, Robert; Frankel, G; Pallen, Mark; Busby, Stephen

DOI:
10.1042/BJ20110912

Biochem. J. (2012) **441**, 247–253 (Printed in Great Britain) doi:10.1042/BJ20110912

online data

Translation of a minigene in the 5′ leader sequence of the enterohaemorrhagic *Escherichia coli LEE1* transcription unit affects expression of the neighbouring downstream gene

Md. Shahidul ISLAM*, Robert K. SHAW*†, Gad FRANKEL†, Mark J. PALLEN* and Stephen J. W. BUSBY*[1]

*School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K., and †Centre for Molecular Microbiology and Infection, Division of Cell and Molecular Biology, Imperial College London, London SW7 2AZ, U.K.

The 5′ end of the major RNA transcript of the *LEE1* operon of enterohaemorrhagic *Escherichia coli* contains ~170 bases before the AUG translation start codon of the first recognized gene, *ler*. This unusually long leader sequence carries three potential alternative AUG start codons. Using a *lac* fusion expression vector, we confirmed that the *ler* gene AUG is functional for translation initiation, and we checked for translation initiation at the three alternative AUG codons. Whereas two of the alternative AUG codons appear incompetent for translation initiation, we detected strong initiation at the third AUG, which is followed by one AAA codon and a UAG stop codon. The location of this very short two-codon open reading frame with respect to the *ler* translation start appears to be critical. Hence mutations that destroy the UAG stop codon, or short deletions between the UAG stop codon and the *ler* translation initiation region, result in big effects on *ler* expression. In the context of the full-length *LEE1* operon leader sequence, translation of this very short two-codon open reading frame is necessary for optimal expression of the *ler* gene and for the subsequent interactions of enterohaemorrhagic *Escherichia coli* with host target cells.

Key words: alternative translation start site, enterohaemorrhagic *Escherichia coli*, leader peptide, *LEE1* regulatory region, mutational analysis, translation initiation.

INTRODUCTION

The first translation initiation signal in many bacterial mRNAs is some distance from the 5′ end and may not involve the first AUG methionine codon [1,2]. It is well established that, for most mRNAs, the primary determinant that assures translation initiation is the Shine–Dalgarno sequence, a four to six base sequence that is complementary to a sequence at the 3′ end of one of the rRNAs [3,4]. A compilation of the 5′ end sequences of *Escherichia coli* mRNAs has shown that most mRNAs have 40–80 untranslated bases, but some are much longer [5]. These sequences, known as leader sequences, can adopt different secondary structures, and it is now known that these structures can play a variety of regulatory roles. Some of the earliest studies of post-transcriptional regulation, which concerned bacterial operons involved in amino acid biosynthesis, established that some leaders encoded a short peptide (known as the leader peptide) whose translation had regulatory consequences on the expression of downstream genes [6–10]. However, aside from these cases, most leaders are assumed to be untranslated, although very few systematic studies have been reported.

In our recent work, we have focused on the expression of the *LEE1* operon of EHEC (enterohaemorrhagic *E. coli*) serotype O157:H7. Recall that EHEC serotype O157:H7 causes haemorrhagic colitis and haemolytic uraemic syndrome in humans, and that the *LEE1* operon controls the expression of major pathogenicity determinants encoded by a pathogenicity island known as the LEE (locus of enterocyte effacement). This control is exerted principally by means of expression of the Ler gene transcription regulatory protein that is encoded by *ler*, the first gene of the *LEE1* transcript [11,12]. In a previous study, we identified the essential elements of the *LEE1* promoter [13], whereas the translation start of Ler had been determined from N-terminal analysis of purified Ler protein [14]. These assignments place the Ler AUG ~170 bases downstream of the 5′ end of the *LEE1* mRNA. This unusually long leader sequence contains mostly adenines and uracils, with three AUG sequences upstream of the Ler AUG start codon. In the present study, we have shown that one of these alternative AUG codons is functional for the initiation of translation of a two-codon ORF (open reading frame). We report that interfering with the translation of this minigene affects Ler expression and the ability of EHEC to interact with human epithelial cells.

EXPERIMENTAL

Bacterial strains, plasmids, promoter fragments and primers

E. coli K-12 strain M182, which carries a deletion of the entire lactose operon [15], was used throughout the present study, and was grown on MacConkey lactose indicator plates, LB (Luria–Bertani) medium or DMEM (Dulbecco's modified Eagle's medium) purchased from Sigma–Aldrich. Some experiments were run in EHEC O157:H7 strains Sakai 813 and EDL933, both lacking the *stx* toxin, obtained from Chihiro Sasakawa (Department of Microbiology and Immunology, Institute of Medial Science, University of Tokyo, Tokyo, Japan) and Arthur Donohue-Rolfe (Department of Biomedical Sciences, Cummings School of Veterinary Medicine, Tufts University, North Grafton, MA, U.S.A.) respectively. The vector plasmids pRW224-0 and pRW225-0 used in the present study for cloning different fragments carrying the *LEE1* regulatory region with or without the *ler* gene translation start are illustrated in Figure 1. Both vector plasmids are RK2-based low-copy-number *lac* expression

Abbreviations used: DMEM, Dulbecco's modified Eagle's medium; EHEC, enterohaemorrhagic *Escherichia coli*; FAS, fluorescent-actin staining; LB, Luria–Bertani; LEE, locus of enterocyte effacement; ORF, open reading frame.

[1] To whom correspondence should be addressed (email s.j.w.busby@bham.ac.uk).

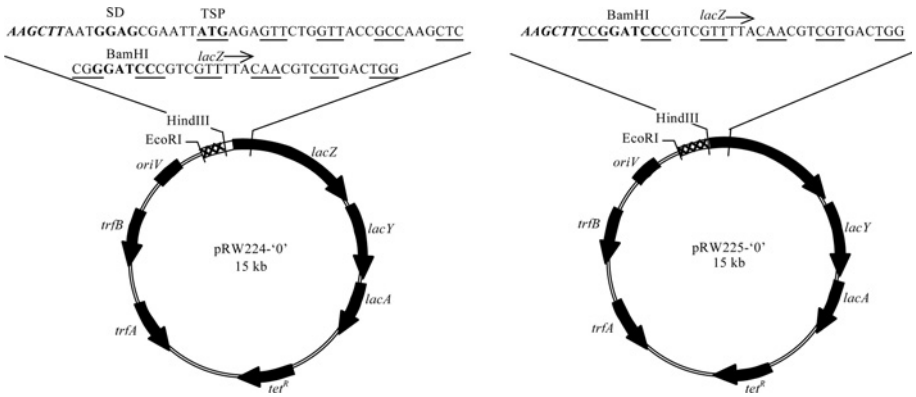

Figure 1 Plasmid maps of the pRW224-0 and pRW225-0 *lac* fusion vectors

The left-hand panel shows an annotated map of plasmid pRW224-0 which was developed as a *lac* expression vector for cloning EcoRI/HindIII promoter fragments that lack a translation start signal. The DNA sequence immediately downstream of the HindIII site is shown and this carries a translation start point (TSP) for the *lacZ* gene. The right-hand panel shows an annotated map of pRW225-0 which was derived from pRW224-0 as a *lac* expression vector for cloning EcoRI/HindIII promoter fragments that carry a translation start signal. This plasmid allows fusion of translation emanating from the fragment to the *lacZ* gene.

vectors, encoding resistance to tetracycline, and were designed to facilitate the cloning of EcoRI/HindIII fragments carrying a promoter directed towards the HindIII end of the fragment [13,16]. Plasmid pRW224-0 carries a translation initiation signal for *lacZ* in the vector immediately downstream of the HindIII site and was designed for cloning promoter fragments that lack a translation start. Plasmid pRW225-0 was derived from pRW224-0 by deleting this translation initiation signal and was designed for cloning promoter fragments that carry a translation start. Hence translation initiation signals that emanate from the cloned sequence are fused to *lacZ* encoded by the pRW225-0 vector (Figure 1).

Standard techniques for recombinant DNA manipulations were used throughout, using PCR with synthetic oligodeoxyribonucleotide primers made by Alta Biosciences that are listed in Supplementary Table S1 (at http://www.BiochemJ.org/bj/441/bj4410247add.htm). All cloned sequences were checked using the Birmingham University Functional Genomics Facility (http://www.genomics.bham.ac.uk/sequencing.htm). PCR was used to amplify the LEE150 fragment using the D61221 and D65811 oligonucleotides and genomic DNA from the O157:H7 Sakai 813 strain. The fragment was made with a flanking EcoRI site upstream of the *LEE1* promoter, and a downstream HindIII site (Figure 2A). The LEE151, LEE30–275 and LEE10–275 fragments were similarly derived by PCR from the LEE150 fragment using the primers listed in Supplementary Table S1.

The different bases of the *LEE1* regulatory region are arbitrarily numbered 1–278 as shown in Figure 2(A). We did not adopt the customary convention of numbering bases with respect to the transcript start point since there is some flexibility in its location, owing to the high AT content of the *LEE1* promoter. Different mutations were introduced into LEE150 and LEE30–275 by megaprimer PCR site-directed mutagenesis [17] using primers listed in Supplementary Table S1. In another experiment, random mutations were introduced into the LEE30–275 fragment by error-prone PCR [18], as described by Islam et al. [19] using flanking primers D63949 and D64101.

Disruption of the *LEE1* mRNA minigene in EHEC and FAS (fluorescent-actin staining) tests

To disrupt the minigene in EHEC strain EDL933, we introduced the 143C and 145G mutations into the EDL933 chromosome using the gene doctoring protocol of Lee et al. [20]. To do this, the LEE150 fragment carrying the 143C and 145G substitutions was cloned into donor plasmid pDOC-C [20] and the recombinant was co-transformed into EDL933 with plasmid pACBSR, which encodes yeast I-SceI meganuclease and the λ-Red recombinase [21]. After transient induction of I-SceI and λ-Red with L-arabinose, recombinants were screened for loss of the donor plasmid and pACBSR, as described in [20], and the transfer of the 143C and 145G mutations to the EDL933 chromosome was checked by sequencing. The FAS method and confocal fluorescence microscopy were used as described by Knutton et al. [22] to measure attachment of EDL933 and the mutant derivative to HeLa cells (kindly provided by Giulio Auciello, School of Biosciences, University of Birmingham).

Promoter activity assays

E. coli strains containing pRW224 or pRW225 carrying different *LEE1* regulatory region fragments were grown as single colonies on MacConkey lactose indicator plates and then inoculated into LB medium supplemented with 35 $\mu g \cdot ml^{-1}$ tetracycline. Cultures were grown aerobically with shaking at 37 °C and harvested during exponential growth, and β-galactosidase expression was measured using the Miller method [23]. Recorded activities are the means for at least three independent experiments and are taken as a measure of the activity of the cloned fragment. In some experiments, LB medium was replaced by DMEM.

RESULTS AND DISCUSSION

Two sites for translation initiation in the *LEE1* operon leader sequence

The starting point for the present study was the LEE150 EcoRI/HindIII DNA fragment, illustrated in Figure 2(A). This fragment covers the EHEC *LEE1* operon regulatory region, including the translation initiation codon that encodes the experimentally determined N-terminal methionine residue of Ler protein, the product of *ler* that is the first recognized gene of the *LEE1* operon. In our previous study [13], we identified the −35 and −10 determinants of the major *LEE1* promoter, P1, and these are highlighted in Figure 2(A). It is apparent that the *LEE1* transcript 5′ leader sequence upstream of the *ler* translation

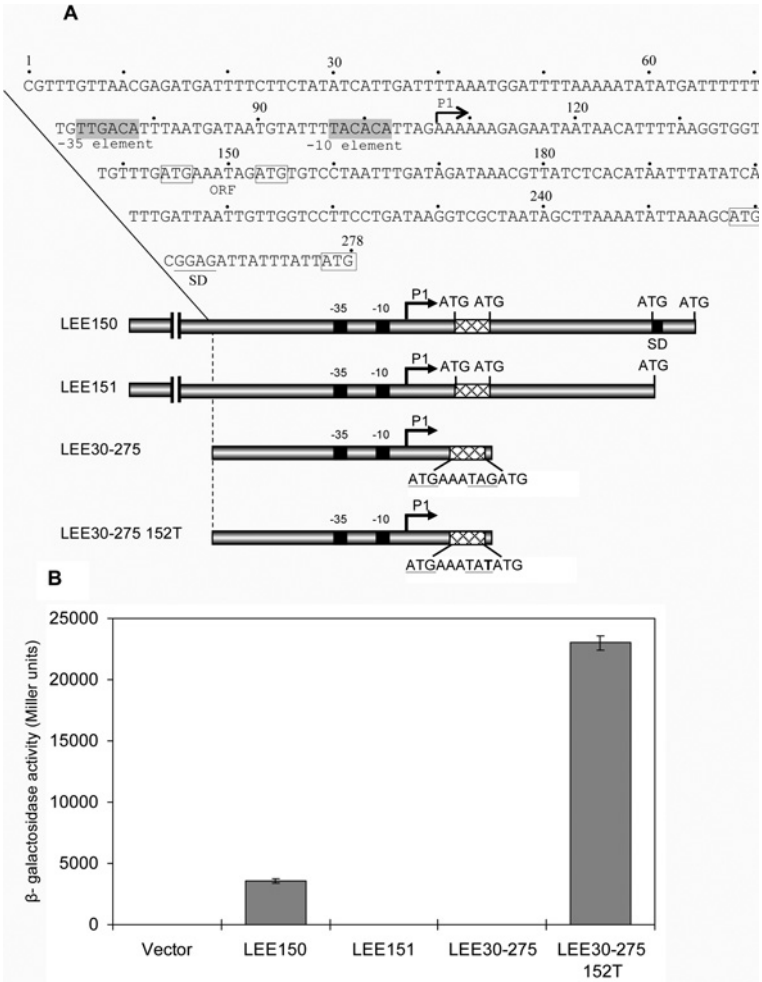

Figure 2 Translation initiation from four AUG codons located in the *LEE1* operon leader region

(**A**) Schematic representation of the LEE150 fragment and three derivatives. The LEE150 fragment covers the *LEE1* operon regulatory region and its partial sequence is shown in the upper part of the Figure. The base sequence is numbered 1–278 starting with the first base located 275 bp upstream of the *ler* translation start codon (position 276–278). The P1 promoter −35 and −10 elements are shaded grey and the transcript start site is indicated by a bent arrow. Methionine triplets in the leader sequence are boxed. The predicted Shine–Dalgarno sequence (SD), which is located in front of the *ler* translation start site, is underlined. Three derivatives (LEE151, LEE30–275 and LEE30–275 152T) of the LEE150 fragment are illustrated with the positions of the promoter elements, the methionine codons and the mini-ORF. (**B**) Measured β-galactosidase activities in the Δ*lac E. coli* K-12 M182 strain carrying empty *lac* expression plasmid pRW225-0 alone (vector) as control, or the plasmid containing either the LEE150 fragment or derivatives. The measurements were made after growing the cells aerobically in LB medium at 37 °C to a D_{650} of ∼0.5. Results are means ± S.D. for at least three independent assays.

initiation codon is ∼170 bases, and inspection of the sequence reveals three other methionine codons. Note that, in the present study and in our previous study [13], the *LEE1* regulatory region is numbered with reference to an arbitrary upstream location that is denoted as position 1 (Figure 2A). Hence the *ler* gene initiator methionine codon is located at co-ordinates 276–278, and the three preceding methionine codons are at positions 144–146, 153–155 and 258–260.

Our aim was to measure translation that starts at the *ler* gene initiator methionine codon and at the three preceding methionine codons. To do this, the LEE150 fragment was cloned into the pRW225 *lac* expression vector so that the *lacZ* gene was fused to the *ler* initiator methionine codon at position 276–278 of the *LEE1* regulatory region. The resulting recombinant plasmid was transformed into the M182 Δ*lac E. coli* strain, and β-galactosidase expression was measured. Results in Figure 2(B) show that cloning the LEE150 fragment into pRW225 leads to accumulation of β-galactosidase in strain M182. Next, the experiment was repeated with the shorter LEE151 fragment (Figure 2A). Cloning the LEE151 fragment into pRW225 fuses the methionine codon at positions 258–260 to the *lacZ* gene. Results in Figure 2(B) show that M182 cells carrying pRW225 with the LEE151 fragment accumulate very little β-galactosidase, with measured levels similar to those found with empty pRW225. Taken together, these data argue that the 276–278, but not the 258–260, methionine codon is associated with a translation initiation signal and confirms the previous identification of the Ler N-terminal amino acid [14].

To investigate translation initiation at methionine codons 144–146 and 153–155, we exploited the LEE30–275 promoter fragment (Figure 2A). Cloning of this fragment into pRW225 fuses the 153–155 methionine codon to the *lacZ* gene. M182 cells carrying this recombinant contain very low levels of β-galactosidase, suggesting that the 153–155 methionine codon is not associated with a translation initiation signal (Figure 2B). Note that our previous results found that the upstream sequences deleted in the LEE30–275 fragment have only a marginal effect on *LEE1* P1 promoter activity [13]. To investigate translation initiation at

Table 1 Effects of mutations on translation of *LEE1* leader mini-ORF

The Table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW225 containing the LEE30–275 fragment with different mutations. Cultures were grown aerobically at 37 °C in LB medium to a D_{650} of ~0.5. Activities were measured in triplicate, giving a mean ± S.D. The top line of the central part of the Table shows the base sequence from positions 141 to 162 of the LEE30–275 fragment that includes the minigene which is shaded grey. LEE30–275 fragments were cloned in pRW225 so that the minigene translation start would be in-frame with the *lacZ* gene, and the sequence of each fusion is shown. Different mutations are listed in the left-hand column and highlighted in bold and underlining in the base sequences.

Derivatives of LEE30–275	Sequence of LEE30–275 fragment from positions 129 to 162 and fusion to *lacZ*	β-Galactosidase activity (Miller units)
Wild-type	141 150 162 TTGATGAAATAGATGTGTCCTA	
Mutations in the translation stop codon		
152T	TTGATGAAATA**T**ATGTGTCCTAAAGCTTGGGATCCCGTCGTT *lacZ*→	22995 ± 561
Mutations in the translation start sites		
145G 152T	TTGA**G**GAAATA**T**ATGTGTCCTAAAGCTTGGGATCCCGTCGTT	2503 ± 17
143C 145G 152T	TT**C**A**G**GAAATA**T**ATGTGTCCTAAAGCTTGGGATCCCGTCGTT	52 ± 2
143C 152T	TT**C**ATGAAATA**T**ATGTGTCCTAAAGCTTGGGATCCCGTCGTT	23375 ± 801

the methionine codon at positions 144–146, we exploited the fact that it is followed by AAA and TAG triplets that are in phase with the 153–155 methionine codon (Figure 2A). Hence the 152T mutation in the LEE30–275 fragment cloned in pRW225 converts the TAG stop triplet to TAT and fuses the 144-146 methionine codon to the *lacZ* gene. M182 cells containing pRW225 with the LEE30–275 fragment carrying the 152T mutation express high levels of β-galactosidase (Figure 2B), suggesting that the 144–146 methionine codon is associated with a functional translation start signal. To test this directly, the 145G mutation was introduced to destroy the 144–146 methionine codon. Results in Table 1 show that, in the context of the LEE30–275 fragment carrying the 152T mutation, the 145G mutation causes a ~90 % reduction in β-galactosidase expression. A similar result was found with the 144T substitution that also destroys the 144–146 methionine codon (Md.S. Islam, unpublished work). Thus, taken together, our results argue that, in the wild-type *LEE1* leader sequence, translation initiates at the 144–146 methionine codon, but must terminate at the TAG stop triplet at position 150–152. Translation of the *ler* gene then starts at the 276–278 methionine codon.

The experiments with pRW225 carrying the LEE30–275 152T fragment, described in Table 1, showed that β-galactosidase expression is not reduced to basal levels by the 145G mutation in the 144–146 methionine codon. Since this codon is preceded by a TTG triplet, and it is known that, in some cases, UUG can be used to initiate translation [2], we constructed a derivative of the LEE30–275 152T fragment carrying the 143C and 145G substitutions, thereby disrupting both potential initiation codons (Table 1). Measurements of β-galactosidase expression show that translation is reduced to basal levels when the 143C and 145G mutations are combined, indicating that translation of the minigene can start at the 141–143 or 144–146 triplets. A further experiment listed in Table 1 showed that, in the context of the LEE30–275 152T fragment, the 143C substitution that disrupts the TTG triplet has little or no effect on expression, suggesting that UUG is used to initiate translation only when the 144–146 AUG methionine codon is mutated.

Translation of the *LEE1* leader minigene affects expression of the next cistron

To investigate possible effects of the *LEE1* leader minigene on expression of downstream genes, we returned to the LEE150 fragment cloned in pRW225 (Figure 2A). Recall that, in this plasmid, the *ler* gene initiator methionine codon at position 276–278 of the *LEE1* regulatory region is fused to *lacZ* and hence β-galactosidase expression measurements can be used as a monitor of *ler* gene expression. Figure 3 illustrates an experiment in which different base changes were used to disrupt the minigene. Introduction of the 143C and 145G substitutions together caused 50 % reduction in expression (Figure 3B), whereas, individually, the two mutations had lesser effects. Similar effects were observed when multiple mutations were introduced to make the LEE150-1 derivative in which the minigene is replaced with a KpnI restriction site. These results argue that minigene expression is needed for optimal *ler* gene expression and, because different base substitutions produce similar changes in expression levels, effects due to alterations in mRNA secondary structure are probably minimal.

In addition to disrupting the minigene in the LEE150 fragment cloned in pRW225, we used the 151T mutation to convert the TAG stop triplet at position 150–152 into a TTG coding triplet. This results in extension of the minigene by 13 codons as ribosomes run to a stop codon at positions 189–191 (Figure 2A). Results illustrated in Figure 3(B) show that this also causes a 50 % reduction in β-galactosidase expression in our assay. Taken together, our results show that minigene translation affects *ler* expression. To check whether our conclusions were biased by the use of an *E. coli* K-12 host strain, the experiments were repeated using the Sakai 813 EHEC O157:H7 strain. Results presented in Figure 3(C), with cells grown in either LB medium or DMEM, show that the effects of the different mutations that affect minigene translation are similar to those observed in the K-12 strain (Figure 3B).

Misplacement of the minigene down-regulates expression of the adjacent gene

The experiment with the 151T substitution described in Figure 3(B) argues that the position of the end of the *LEE1* leader minigene affects expression of the adjacent downstream gene. To investigate this further, we exploited the expression vector plasmid pRW224 that carries a translation initiation signal for *lacZ* in the vector immediately downstream of the HindIII site (Figure 1), and the LEE30–275 and LEE10–275 EcoRI/HindIII fragments, covering the *LEE1* regulatory region, illustrated in Figure 4(A), were cloned into this vector. Both fragments carry the *LEE1* P1 promoter, and, in the LEE30–275 fragment, the HindIII site is located immediately downstream of the stop codon of the leader

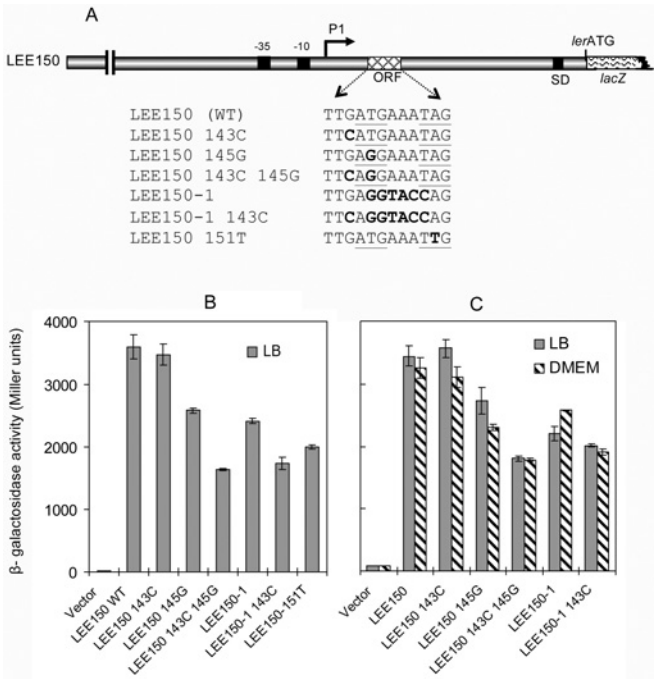

Figure 3 Effects of translation of the *LEE1* leader minigene on the expression of the downstream gene

(**A**) Schematic representation of the LEE150 fragment, and derivatives carrying mutations in the leader minigene. LEE150 145G carries a point mutation in the ATG start triplet of the minigene. LEE150 143C carries a point mutation in the upstream TTG triplet. LEE150 143C 145G carries mutations that altered both triplets. In LEE150-1, the predicted minigene was replaced with a KpnI site. LEE150-1 143C carries a mutation in the second translation start site. In LEE150 151T, the translation stop triplet, TAG, was changed to TTG. Derivatives were cloned as translation fusions to *lacZ* into pRW225. (**B**) Measured β-galactosidase activities in Δ*lac E. coli* K-12 strain M182 containing pRW225 carrying LEE150 or derivatives. Measurements were made after growing the cells aerobically in LB medium at 37 °C to a D_{650} of ~0.5. (**C**) Measured β-galactosidase activities in EHEC O157:H7 Sakai 813 cells. Measurements were made after growing the cells aerobically either in LB medium (grey bars) or in DMEM (striped bars) at 37 °C to a D_{650} of ~0.5. Results are means ± S.D. for at least three independent assays.

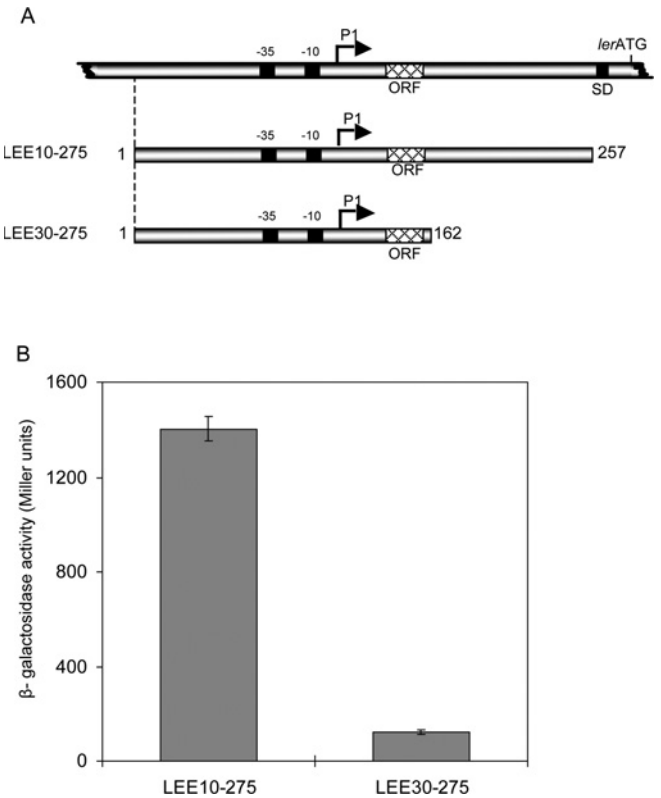

Figure 4 Expression from different *LEE1* promoter regulatory region fragments

(**A**) Schematic representation of the LEE10–275 and LEE30–275 fragments that carry the *LEE1* operon regulatory region. The upstream and downstream end of each derivative is numbered following the numbering system in Figure 2. The locations of the P1 promoter functional elements and the minigene are indicated. (**B**) Measured β-galactosidase activities in Δ*lac E. coli* K-12 strain M182 containing the plasmid pRW224 with either LEE10–275 or LEE30–275 fragments as transcriptional fusions to *lacZ*. Measurements were made after growing the cells aerobically in LB medium at 37 °C to a D_{650} of ~0.5. Results are means ± S.D. for at least three independent assays.

minigene. Hence, when cloned in pRW224, the spacing between the minigene stop codon and the *lacZ* translation initiation codon is 29 bp. In contrast, when the LEE10–275 EcoRI/HindIII fragment was cloned into pRW224, the spacing between the minigene stop codon and the *lacZ* translation initiation codon is 124 bp.

Measurements of β-galactosidase expression in M182 cells containing pRW224 carrying either the LEE10–275 or LEE30–275 fragments show that, even though both constructions carry the same promoter and the same translation start region for the *lacZ* gene, expression levels differ ~10-fold. In particular, with pRW224 carrying the LEE30–275 fragment, expression was unexpectedly low. To identify factors responsible for this low level of β-galactosidase expression, we set up a screen to find mutations in the LEE30–275 fragment that resulted in higher expression levels. To do this, error-prone PCR was used to generate six independent preparations of the LEE30–275 fragment carrying randomly generated mutations. The fragments were then cloned into pRW224, the resulting recombinant plasmids were transformed into *E. coli* strain M182, and transformants were grown on MacConkey lactose indicator plates. As expected, the vast majority of colonies scored as Lac^- (white), but, after screening over 10000 transformants, we identified ~36 Lac^+ (pink) colonies. DNA sequencing showed that half of these carried single base substitutions, and hence we identified 13 different single base substitutions that increased β-galactosidase expression from pRW224 carrying the LEE30–275 fragment. Their locations are shown in Figure 5, and different measured β-galactosidase activities are presented in Table 2. The mutations fall in clusters. One cluster (99T, 101A, 101T and 102T) falls in the P1 − 10 hexamer (TACACA) and the effects of these mutations are probably due to increased transcription. Since the 102T mutation changes the hexamer towards the consensus, TATAAT, it is likely to increase *LEE1* P1 promoter activity. Similarly, the 99T, 101A and 101T base changes remove non-consensus C residues from the hexamer, and so also increase promoter activity. A second cluster of mutations (132A, 132T, 133A, 135T and 144T) falls in the translation initiation region of the *LEE1* leader minigene. The 132A, 132T, 133A and 135T mutations corrupt the Shine–Dalgarno sequence, and the 144T change falls in the initiator methionine triplet. Since the mutations in this cluster will reduce translation of the minigene, we conclude that, with pRW224 carrying the LEE30–275 fragment, expression of the minigene must be detrimental to expression of the adjacent *lacZ* gene. Of the four other mutations, two (116T and 122C) are located towards the 5′ end of the *LEE1* transcript and no simple explanation for their effects is apparent. In contrast, Δ150 and 151T destroy the stop codon of the minigene and hence alter the juxtaposition of its translation stop with respect to the *lacZ* gene.

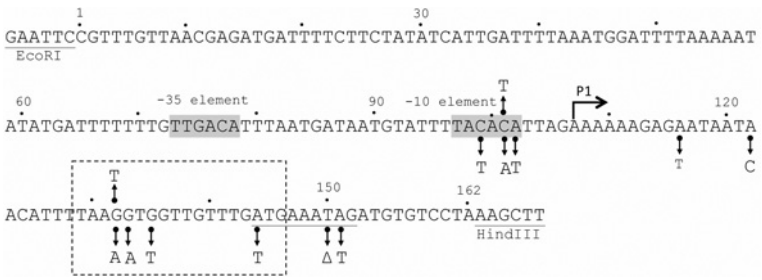

Figure 5 Mutational analysis of the LEE30–275 fragment

Nucleotide sequence of the upper strand of the EcoRI/HindIII LEE30–275 fragment. The base sequence is numbered 1–162 as in Figure 2. The P1 promoter −35 and −10 hexamer elements are shaded grey and the transcript start point is indicated by a bent arrow. The positions of randomly generated mutations that increased expression from this fragment when cloned in pRW224 are illustrated by showing the substituted bases with vertical arrows. The broken box represents the predicted minigene translation initiation region, whereas the bases of the minigene are underlined.

Table 2 Mutational analysis of the LEE30–275 promoter fragment

The Table shows measured β-galactosidase activities in cultures of *E. coli* strain M182 carrying pRW224 with the LEE30–275 fragment and different mutant derivatives. Measurements were made after growing the cells aerobically in LB medium at 37 °C to a D_{650} of ~0.5 and are listed in the central column. Activities were measured in triplicate, giving a mean ± S.D. Activities expressed as fold increases compared with the starting LEE30–275 fragment are given in the third column.

Promoter fragments	β-Galactosidase activity (Miller units)	Fold increase
Starting fragment		
LEE30–275	124 ± 9	–
Mutations at P1 promoter −10 element		
LEE30–275 99T	554 ± 18	4.5
LEE30–275 101A	565 ± 18	4.6
LEE30–275 101T	412 ± 9	3.3
LEE30–275 102T	928 ± 31	7.5
Mutations in the minigene translation initiation region		
LEE30–275 132A	551 ± 28	4.5
LEE30–275 132T	498 ± 130	4.0
LEE30–275 133A	460 ± 9	3.7
LEE30–275 135T	401 ± 33	3.2
LEE30–275 144T	288 ± 11	2.3
Mutations at other positions		
LEE30–275 116T	765 ± 53	6.2
LEE30–275 122C	948 ± 126	7.6
LEE30–275 Δ150	357 ± 13	2.9
LEE30–275 151T	1270 ± 109	10.2

Role of the EHEC *LEE1* operon minigene *in vivo*

Our principal conclusion is that the long leader sequence of the *LEE1* operon of EHEC serotype O157:H7 encodes a two-codon minigene which is translated. Its translation has a positive effect *in cis* on expression of the adjacent *ler* gene, which encodes a transcription factor that orchestrates the expression of key pathogenicity determinants of EHEC. Our experiments show that changing the distance between the minigene and the adjacent downstream gene can result in minigene translation being inhibitory to expression of the downstream gene. Although the molecular reasons for these effects are as yet undefined, it is not hard to see how minigene translation could affect leader sequence conformation or stability, and that this could depend on many factors including the precise position and length of the minigene. In order to investigate the importance of the minigene in EHEC, we engineered a derivative of the EHEC EDL933 strain carrying the 143C and 145G mutations that completely suppress minigene translation and used the well-established FAS test [22] to investigate the effects of the mutations on interactions of

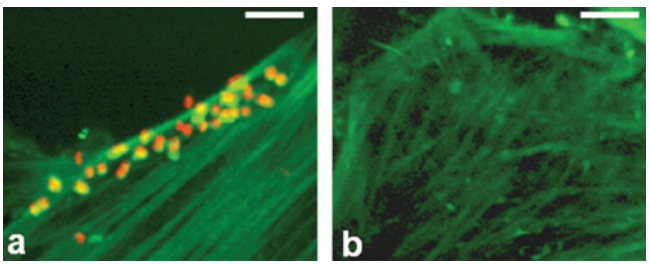

Figure 6 Translation of the *LEE1* leader minigene is essential for formation of attaching and effacing lesions

EHEC strain EDL933 interacts with HeLa cells detected using the FAS method and confocal fluorescence microscopy. Bacterial are identified by red propidium iodide staining of their chromosomal DNA, and actin appears green. Attachment of EHEC to HeLa cells results in the formation of actin foci that appear as intense yellow dots. (**a**) Experiment run with wild-type EHEC strain EDL933. (**b**) The same experiment, but with the mutant strain carrying the 143C and 145G mutations that stop translation of the *LEE1* operon leader minigene. Scale bars, 5 μm.

EDL933 with human epithelial cells. Recall that, upon attachment to epithelial cells, EHEC uses a LEE-encoded type 3 secretion system to inject effector molecules that results in the so-called 'attaching and effacing' lesion and the accumulation of actin at the site of attachment (for a recent review, see [24]). In the FAS test, phalloidin labelled with FITC is used to detect this accumulation, using fluorescence microscopy [22]. Figure 6(a) shows clear foci of actin accumulation upon interaction of the starting EHEC EDL933 strain with HeLa cells. In contrast with the EDL933 derivative in which minigene translation is suppressed, no foci are seen and there is no bacterial attachment to the HeLa cells (Figure 6b). Since this phenotype is not complemented by a plasmid carrying the LEE150 fragment (Md. S. Islam, unpublished work), we conclude that the minigene is *cis*-acting rather than *trans*-acting, and that the reduction in Ler expression due to suppression of minigene translation results in an EHEC strain that is unable to interact normally with epithelial cells.

Similar minigenes were reported by Guarneros and colleagues [25,26] in the leader sequences of some bacteriophage λ mRNAs, but there are very few other characterized examples and their significance is unclear. Remarkably, the *LEE1* leader minigene is highly conserved in all pathogenic *E. coli* strains containing a LEE pathogenicity island [27], and in *Citrobacter rodentium* [28]. In all cases, the minigene contains two adjacent possible translation start codons and a sense codon specific for lysine (see Supplementary Figure S1 at http://www.BiochemJ.org/bj/441/bj4410247add.htm). Our results suggest that the *LEE1* leader minigene may have evolved as an intricate device to set levels of *ler* gene expression.

Finally, for many bacterial mRNAs, ribosomes have a choice of different potential translation start codons, and, in the case of the *LEE1* leader sequence, there appear to be four possible choices. Interestingly, the widely used RBS (ribosome binding sites) scoring programme [29] predicts translation initiation at the methionine triplets at positions 144–146 and 276–278, but not at positions 153–155 and 258–260, suggesting that translation initiation preferences follow the conventional rules. However, as yet, we are unable to account for the key role of spacing in setting the level of expression of the gene located downstream of the minigene.

AUTHOR CONTRIBUTION

The work was performed by Md. Shahidul Islam as part of his doctoral work supervised by Stephen Busby and Mark Pallen. Work with EHEC strains was part of a collaboration with Robert Shaw and Gad Frankel.

ACKNOWLEDGEMENTS

We are grateful to Dave Lee for help with gene doctoring and Chris Webster for technical support.

FUNDING

This work was supported by a Biotechnology and Biological Sciences Research Council project grant, a Wellcome Trust Programme Grant, and a Commonwealth Commission PhD studentship to M.S.I.

REFERENCES

1 Kozak, M. (1983) Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**, 1–455
2 O'Donnell, S. M. and Janssen, G. R. (2001) The initiation codon affects ribosome binding and translational efficiency in *Escherichia coli* of cI mRNA with or without the 5′ untranslated leader. J. Bacteriol. **183**, 1277–1283
3 Shine, J. and Dalgarno, L. (1974) The 3′-terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. U.S.A. **71**, 1342–1346
4 Steitz, J. A. and Jakes, K. (1975) How ribosomes select initiator regions in mRNA: base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. Proc. Natl. Acad. Sci. U.S.A. **72**, 4734–4738
5 Shultzaberger, R. K., Chen, Z., Lewis, K. A. and Schneider, T. D. (2007) Anatomy of *Escherichia coli* σ^{70} promoters. Nucleic Acids Res. **35**, 771–788
6 Gardner, J. F. (1979) Regulation of the threonine operon: tandem threonine and isoleucine codons in the control region and translational control of transcription termination. Proc. Natl. Acad. Sci. U.S.A. **76**, 1706–1710
7 Gemmill, R. M., Wessler, S. R., Keller, E. B. and Calvo, J. M. (1979) *leu* operon of *Salmonella typhimurium* is controlled by an attenuation mechanism. Proc. Natl. Acad. Sci. U.S.A. **76**, 4941–4945
8 Johnston, H. M., Barnes, W. M., Chumley, F. G., Bossi, L. and Roth, J. R. (1980) Model for regulation of the histidine operon of *Salmonella*. Proc. Natl. Acad. Sci. U.S.A. **77**, 508–512
9 Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. Nature **289**, 751–758
10 Gavini, N. and Pulakat, L. (1991) Role of translation of the *pheA* leader peptide coding region in attenuation regulation of the *Escherichia coli pheA* gene. J. Bacteriol. **173**, 4904–4907
11 Perna, N. T., Mayhew, G. F., Pósfai, G., Elliott, S., Donnenberg, M. S., Kaper, J. B. and Blattner, F. R. (1998) Molecular evolution of a pathogenicity island from enterohemorrhagic *Escherichia coli* O157:H7. Infect. Immun. **66**, 3810–3817
12 Mellies, J. L., Elliott, S. J., Sperandio, V., Donnenberg, M. S. and Kaper, J. B. (1999) The *Per* regulon of enteropathogenic *Escherichia coli*: identification of a regulatory cascade and a novel transcriptional activator, the locus of enterocyte effacement (LEE)-encoded regulator (Ler). Mol. Microbiol. **33**, 296–306
13 Islam, M. S., Bingle, L. E. H., Pallen, M. J. and Busby, S. J. W. (2010) Organization of the *LEE1* operon regulatory region of enterohaemorrhagic *Escherichia coli* O157:H7 and activation by GrlA. Mol. Microbiol. **79**, 468–483
14 Yerushalmi, G., Nadler, C., Berdichevski, T. and Rosenshine, I. (2008) Mutational analysis of the locus of enterocyte effacement-encoded regulator (*Ler*) of enteropathogenic *Escherichia coli*. J. Bacteriol. **190**, 7808–7818
15 Casadaban, M. J. (1976) Transposition and fusion of the *lac* genes to selected promoters in *Escherichia coli* using bacteriophage lambda and Mu. J. Mol. Biol. **5**, 541–555
16 Lodge, J., Fear, J., Busby, S., Gunasekaran, P. and Kamini, N.-R. (1992) Broad host range plasmids carrying the *Escherichia coli* lactose and galactose operons. FEMS Microbiol. Lett. **95**, 271–276
17 Sarkar, G. and Sommer, S. S. (1990) The "megaprimer" method of site directed mutagenesis. Biotechniques **8**, 404–407
18 Leung, D. W., Chen, E. and Goeddel, D. V. (1989) A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction. Technique **1**, 11–15
19 Islam, M. S., Pallen, M. J. and Busby, S. J. (2011) A cryptic promoter in the *LEE1* regulatory region of enterohaemorrhagic *Escherichia coli*: promoter specificity in AT-rich gene regulatory regions. Biochem. J. **436**, 681–686
20 Lee, D. J., Bingle, L. E., Heurlier, K., Pallen, M. J., Penn, C. W., Busby, S. J. and Hobman, J. L. (2009) Gene doctoring: a method for recombineering in laboratory and pathogenic *Escherichia coli* strains. BMC Microbiol. **9**, 252
21 Herring, C. D., Glasner, J. D. and Blattner, F. R. (2003) Gene replacement without selection: regulated suppression of amber mutations in *Escherichia coli*. Gene **311**, 153–163
22 Knutton, S., Baldwin, T., Williams, P. H. and McNeish, A. S. (1989) Actin accumulation at sites of bacterial adhesion to tissue culture cells: basis of a new diagnostic test for enteropathogenic and enterohemorrhagic *Escherichia coli*. Infect. Immun. **57**, 1290–1298
23 Miller, J. (1972), Experiments in Molecular Genetics, Cold Spring Harbor Laboratory Press, Cold Spring Harbor
24 Wong, A. R. C., Pearson, J. S., Bright, M. D., Munera, D., Robinson, K. S., Lee, S. F., Frankel, G. and Hartland, E. L. (2011) Enteropathogenic and enterohemorrhagic *Escherichia coli*: even more subversive elements. Mol. Microbiol. **80**, 1420–1438
25 Ontiveros, C., Valadez, J. G., Hernández, J. and Guarneros, G. (1997) Inhibition of *Escherichia coli* protein synthesis by abortive translation of phage lambda minigenes. J. Mol. Biol. **6**, 167–175
26 Oviedo, A.-N., Salgado, H., Collado-Vides, J. and Guarneros, G. (2004) Distribution of minigenes in the bacteriophage lambda chromosome. Gene **329**, 115–124
27 Chaudhuri, R. R., Loman, N. J., Snyder, L. A. S., Bailey, C. M., Stekel, D. J. and Pallen, M. J. (2008) xBASE2: a comprehensive resource for comparative bacterial genomics. Nucleic Acids Res. **36**, D543–D546
28 Deng, W., Li, Y., Vallance, B. A. and Finlay, B. B. (2001) Locus of enterocyte effacement from *Citrobacter rodentium*: sequence analysis and evidence for horizontal transfer among attaching and effacing pathogens. Infect. Immun. **69**, 6323–6335
29 Salis, H., Mirsky, E. A. and Voigt, C. A. (2009) Automated design of synthetic ribosome binding sites to control protein expression. Nat. Biotechnol. **27**, 946–950

Received 23 May 2011/5 October 2011; accepted 6 October 2011
Published as BJ Immediate Publication 6 October 2011, doi:10.1042/BJ20110912

Biochem. J. (2012) **441**, 247–253 (Printed in Great Britain) doi:10.1042/BJ20110912

online data

SUPPLEMENTARY ONLINE DATA

Translation of a minigene in the 5′ leader sequence of the enterohaemorrhagic *Escherichia coli LEE1* transcription unit affects expression of the neighbouring downstream gene

Md. Shahidul ISLAM*, Robert K. SHAW*†, Gad FRANKEL†, Mark J. PALLEN* and Stephen J. W. BUSBY*[1]

*School of Biosciences, University of Birmingham, Edgbaston, Birmingham B15 2TT, U.K., and †Centre for Molecular Microbiology and Infection, Division of Cell and Molecular Biology, Imperial College London, London SW7 2AZ, U.K.

Table S1 Synthetic oligodeoxynucleotides used in the present study

Restriction sites are underlined.

Name	Sequence (5′→3′)	Use
D61221	GCAGAATTCTGCACCCGTTCCAGG	Upstream primer containing EcoRI site for amplification of EcoRI/HindIII LEE150 fragment. Cloned into pRW225-0
D65811	GCAAAGCTTCATAATAAATAATCTCCG	Downstream primer containing HindIII site used in PCR with D61221 for amplification of LEE150. Cloned into pRW225-0
D65812	GCAAAGCTTCATGCTTTAATATTTTAAGCT	Downstream primer containing HindIII site used in PCR with D61221 for amplification of LEE151.Cloned into pRW225-0
D63949	GCAGAATTCCGTTTGTTAACGAGATGATTTTCTTC	Upstream primer for amplification of EcoRI/HindIII LEE30–275 fragment. Cloned into pRW224-0
D64101	GCAAAGCTTTAGGACACATCTATTTCA	Downstream primer site used in PCR with upstream D63949 primer to produce EcoRI/HindIII LEE30–275 fragment. Cloned into pRW224-0
D65474	GCAGGATCCCAAGCTTTAGGACACATC	Downstream primer used in PCR with D63949 to amplify EcoRI/BamHI LEE30–275 fragment. Cloned into pRW225-0
D66870	GAAATATATGTGTCCTAAAGCTT	Upstream primer used in PCR with D53463 to produce megaprimer for construction of LEE30–275 152T fragment
D61222	GCAAAGCTTGCTTTAATATTTTAAGC	Downstream primer containing HindIII site used in PCR with D63949 for amplification of EcoRI/HindIII LEE10–275 fragment. Cloned into pRW224-0
D10520	CCCTGCGGTGCCCCTCAAC	Anneals upstream of EcoRI site in pRW224-0 or pRW225-0. Used for sequencing cloned fragments
D53463	GGGGGATGTGCTGCAAGGCG	Anneals downstream of HindIII site in pRW224 or pRW225-0. Used for sequencing cloned fragments
D67046	GTGGTTGTTTCATGAAATAGATG	Upstream primer used in PCR with D65811 to produce megaprimer for construction of LEE150 143C
D66271	GTGGTTGTTTGAGGAAATAGATGTGT	Upstream primer used in PCR together with D65811 to produce megaprimer for construction of LEE150 145G
D66270	TAAGGTGGTTGTTTCAGGAAATAGATGTGTC	Upstream primer used in PCR with D65811 to produce megaprimer for construction of LEE150 143C 145G
D66032	GCAGGTACCAGATGTGTCCTAATTTGATAG	Upstream primer used in PCR with D65811 to generate downstream KpnI fragment for construction of LEE150-1 fragment
D66033	GCAGGTACCTCAAACAACCACCTTAAAATG	Downstream primer used in PCR with D61221 to generate upstream KpnI fragment for construction of LEE150-1 fragment
D66176	TCTGGTACCTGAAACAACCAC	Downstream primer used in PCR with D61221 to produce a megaprimer for construction of LEE150-1 143C fragment

Figure S1 Alignment of partial DNA sequences of the *LEE1* operon regulatory region of different strains of *E. coli* and *C. rodentium*

Multiple sequence alignment was performed using ClustalW [1]. Promoter −35 and −10 elements are indicated by white boxes. The light grey box corresponds to predicted mini-ORFs, which have potential Shine–Dalgarno sequences (SD) (shaded dark grey) preceding them. Ec and Cr refer to *E. coli* and *C. rodentium* respectively.

[1] To whom correspondence should be addressed (email s.j.w.busby@bham.ac.uk).

REFERENCE

1 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignments through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**, 4673–4680

Received 23 May 2011/5 October 2011; accepted 6 October 2011
Published as BJ Immediate Publication 6 October 2011, doi:10.1042/BJ20110912